Creation of a Dataset for the first step of STPA and Machine Learning Classification

Andrey Toshiro Okamura
Universidade Estadual de Campinas
a213119@dac.unicamp.br

Ana Estela Antunes da Silva
Universidade Estadual de Campinas
aeasilva@unicamp.br

Luiz Eduardo Galvão Martins
Universidade Federal de São Paulo
legmartins@unifesp.br

AGENDA

- Introduction
- Methodology
- Results and Discussion
- Future work

AGENDA

- **Introduction**
- Methodology
- Results and Discussion
- Future work

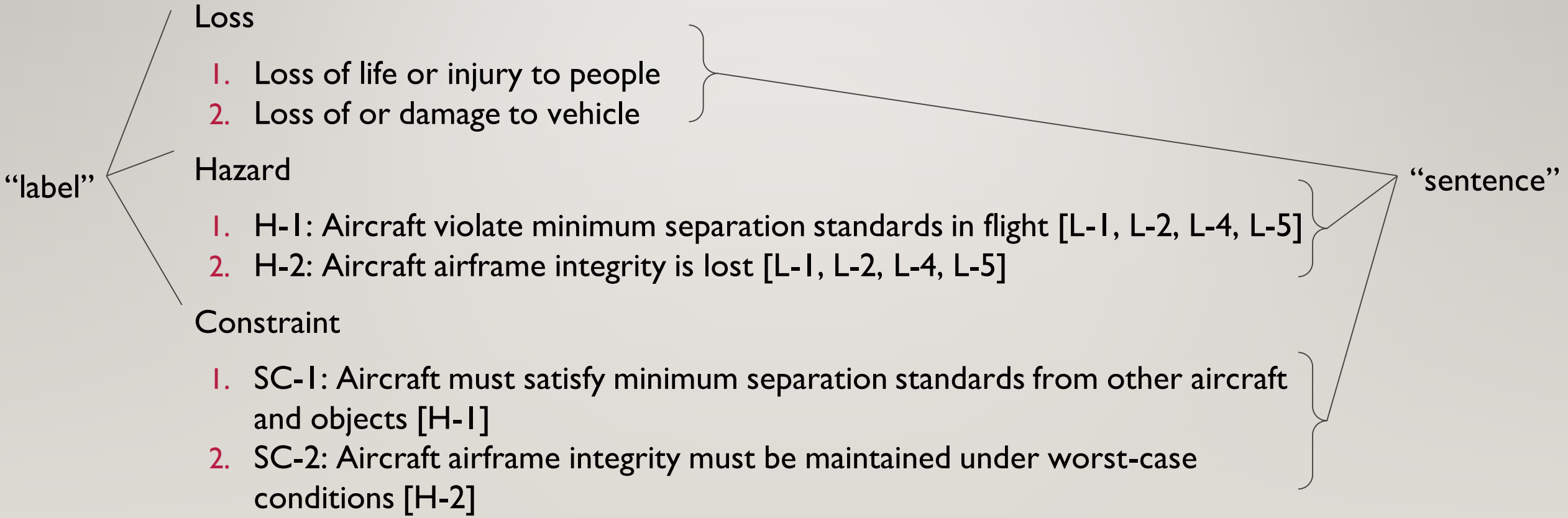
4 Introduction

Increase in use of Machine learning (ML) and Natural Language Processing (NLP).

No datasets related to STPA and ML is publicly available.

Start by creating a dataset for the first step of STPA ("Defining the purpose of the Analysis").

5 Introduction



Sentence: Loss of life or injury to people	**Label:** Loss

AGENDA

- Introduction
- **Methodology**
- Results and Discussion
- Future work

7 Methodology (dataset creation)

Extraction of sentences from a publicly available source (MIT STAMP Workshop);

Presentations are from 2012-2023;

For each year, all available presentations were opened. A manual search for lists or tables containing examples of the first STPA step was performed.

The extracted sentences were recorded in a spreadsheet, along with their labels and metadata.

https://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/

8 Methodology (dataset creation)

1: "sentence": The extracted textual sentence;

2: "label": The classification label related to the sentence;

3: "domain": The presentation domain;

4: "year": The year of presentation;

5: "title": The presentation title;

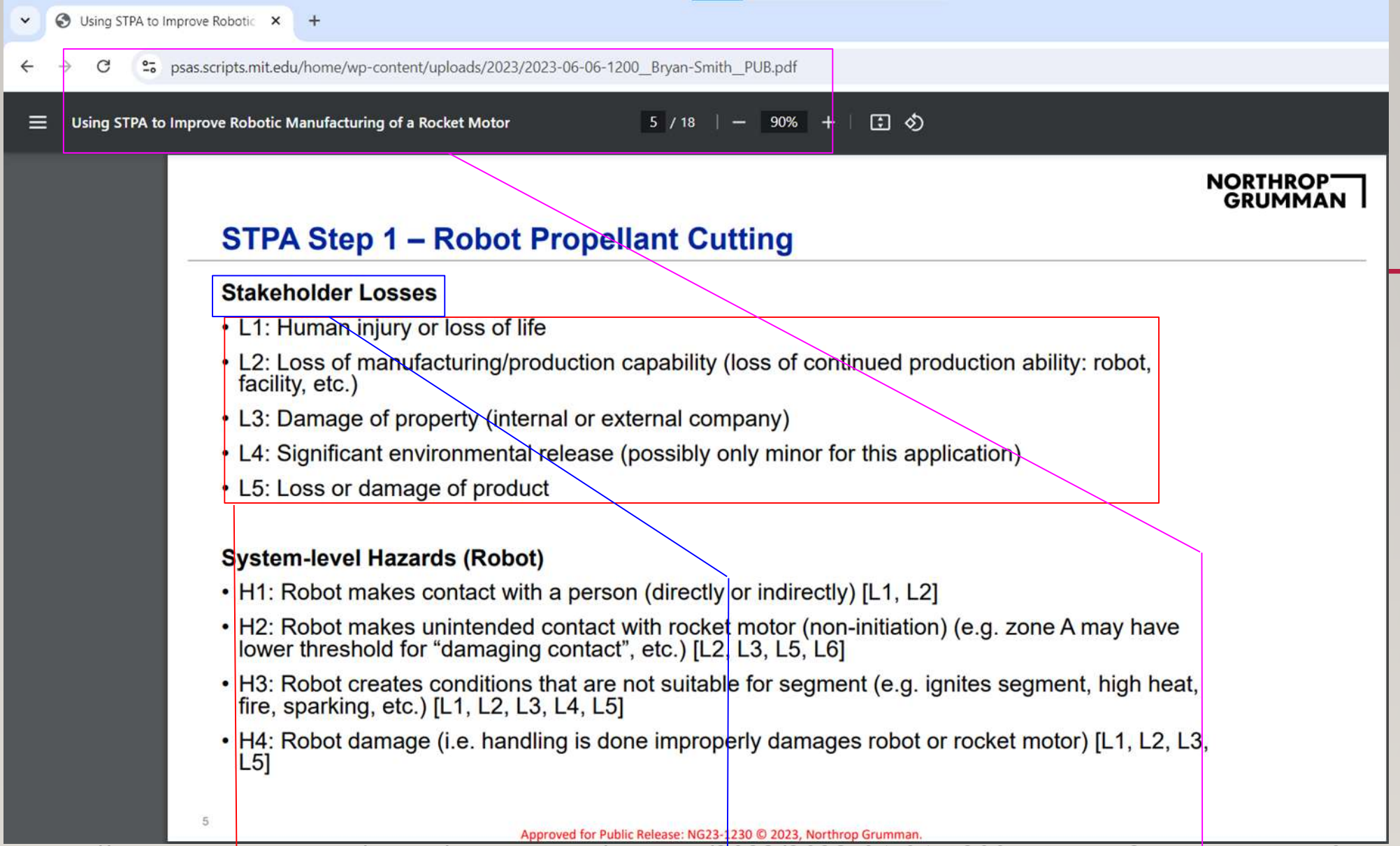
6: "url": The presentation URL;

7: "slide": The slide number where the sentence was extracted;

8: "obs": If the presentation is not explicitly about STPA, then the type of presentation.

9 Methodology (dataset creation)

	A	B	C	D	E	F	G	H
1	**sentence**	**label**	**domain**	**year**	**title**	**url**	**slide**	**obs**
2	Humans and/or human assets on earth are killed/damaged.	loss	aerospace	2012	STPA and CAST	http://psas.script	13	
3	Humans and/or human assets off of the earth are killed/damaged.	loss	aerospace	2012	STPA and CAST	http://psas.script	13	
4	Organisms on any of the moons of the outer planet (if they exist) are killed or mutated by	loss	aerospace	2012	STPA and CAST	http://psas.script	13	
5	The scientific data corresponding to the mission goals are not collected.	loss	aerospace	2012	STPA and CAST	http://psas.script	13	
6	The scientific data is rendered unusable (e.g., deleted, corrupted, not returned at require	loss	aerospace	2012	STPA and CAST	http://psas.script	14	
7	Organisms of Earth origin are mistaken for organisms indigenous to any of the moons of	exloss	aerospace	2012	STPA and CAST	http://psas.script	14	
8	An incident during this mission directly causes another mission to fail to collect, return, a	exloss	aerospace	2012	STPA and CAST	http://psas.script	14	
9	Inability of Mission to collect data.	hazard	aerospace	2012	STPA and CAST	http://psas.script	20	
10	Inability of Mission to return collected data.	hazard	aerospace	2012	STPA and CAST	http://psas.script	20	
11	Inability of Mission scientific investigators to use returned data.	hazard	aerospace	2012	STPA and CAST	http://psas.script	20	
12	Contamination of Outer Planet Moon with biological agents of Earth origin on mission ha	exhazard	aerospace	2012	STPA and CAST	http://psas.script	20	
13	Exposure of Earth life or human assets on Earth to toxic, radioactive, or energetic eleme	exhazard	aerospace	2012	STPA and CAST	http://psas.script	20	
14	Exposure of Earth life or human assets off Earth to toxic, radioactive, or energetic eleme	exhazard	aerospace	2012	STPA and CAST	http://psas.script	20	
15	Inability of other space exploration missions to use shared space exploration infrastructu	hazard	aerospace	2012	STPA and CAST	http://psas.script	20	
16	Near mid-air collision (NMAC): Two controlled aircraft violate minimum separation standa	hazard	aerospace	2012	STPA and CAST	http://psas.script	21	
17	Controlled maneuver into ground.	hazard	aerospace	2012	STPA and CAST	http://psas.script	21	
18	Pilot loses control of aircraft.	exhazard	aerospace	2012	STPA and CAST	http://psas.script	21	
19	Interference with other safety-related aircraft systems.	hazard	aerospace	2012	STPA and CAST	http://psas.script	21	



https://psas.scripts.mit.edu/home/wp-content/uploads/2023/2023-06-06-1200__Bryan-Smith__PUB.pdf

sentence	label	domain	year	title	url	slide	obs
Humans and/or human assets on earth are killed/damaged.	loss	aerospace	2012	STPA and CAST	http://psas.script	13	
Humans and/or human assets off of the earth are killed/damaged.	loss	aerospace	2012	STPA and CAST	http://psas.script	13	

11 Methodology (dataset creation)

Sentence selection:

In this dataset, there are few sentences with ambiguous meanings, or lack of information or context. Some sentences might be too different from what is recommended by the STPA Handbook [1].

Instead of completely removing from the dataset, the sentences which may impact classification performance were grouped into a new label, named “excluded” sentences.

In order to keep the original label information, a combination of both labels is used.

“excluded” losses -> “ex”+”loss” -> “exloss”

Methodology (dataset creation)

Inclusion criteria:

For Losses:
1. The sentence should contain a Loss-related keyword (such as “loss”, “damage”, “injury”);
2. The sentence should involve something of value to stakeholders.

For Hazards:
1. The sentence should mention a <system> and an <unsafe condition>;
2. The sentence should be a state or condition that, together with a set of worst-case environmental conditions, will lead to a loss.

For Constraints:
1. The sentence should mention a <system> and <condition to enforce> (using a modal verb, such as “must”, ”shall”, “should”);
2. The sentence can also define how to minimize losses in case a hazard occurs.

13 Methodology (classification experiments)

Two experiments were carried out in Python programming language to demonstrate the use of this dataset. Both experiments use two traditional machine learning classification algorithms, called Support Vector Machines (SVM) and Naïve Bayes (NB). Both commonly used for NLP in Requirements Engineering [2].

Experiment 1: Classification of sentences only using the "loss", "hazard" and "constraint" labels, without the "excluded" sentences. This aims to investigate the dataset in a clean state, with sentences closer to what is recommended by the Handbook.

Experiment 2: Classification of sentences with the "excluded" sentences reverted and added back into their original classes ("exloss", "exhazard" and "exconstraint" added back into "loss", "hazard" and "constraint", respectively). This aims to investigate the dataset by including possible noise from excluded sentences, and to compare the dataset in different levels of quality.

AGENDA

- Introduction
- Methodology
- **Results and Discussion**
- Future work

15 Results and Discussion

This dataset contains a total of 1078 sentences.

Unbalanced dataset with high hazard occurrence.

Loss: 27.9%;

Hazard: 41.0%;

Constraint: 31.1%;

The domain with most occurrence is aviation, around 31.6% of all sentences.

The dataset is available at GitHub [3].

labels	sentences
loss	301
hazard	442
constraint	335
total	**1078**

labels	sentences
loss	254
exloss	47
hazard	408
exhazard	34
constraint	316
exconstraint	19
total	**978**
total “ex”	**100**

16 Results and Discussion

- Experiment 1, which represents the dataset in a cleaner state, resulted in the best accuracy (and other metrics) for both classifiers. The highest accuracy achieved is 95.40% by SVM, followed by NB with 85.71% accuracy.
- In Experiment 2, which represents the dataset without any filtering of sentences, showed inferior results with an accuracy of 92.59% by SVM and 80.00% by NB.

Experi-ment	Algo-rithm	Accu-racy	Preci-sion	Recall	F-score
1	SVM NB	0.9540 0.8571	0.9670 0.8710	0.9424 0.8639	0.9516 0.8629
2	SVM NB	0.9259 0.8009	0.9334 0.8099	0.9236 0.8054	0.9277 0.8075

17 Results and Discussion

Certain classes are easier to be identified, compared to others.

The constraint class, for example, is well-defined and sentences are similar to each other. All constraints should contain a modal verb such as "must" and "should", which facilitates the classification.

The same happens to the loss class, in which words like "loss of" or "damage to" are frequent.

However, losses that do not use any of those keywords explicitly are susceptible to being mistaken by hazards, which amongst the three classes, has the least defined characteristic for the class.

AGENDA

- Introduction
- Methodology
- Results and Discussion
- **Future work**

19 Future work

Expand the dataset with sentences from new sources;

Verify rules and sentences with specialists;

(New version in development.)

20 References

[1] LEVESON, N.; THOMAS, J. STPA Handbook. 2018. Available at: https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf.

[2] Zamani, Kareshna, et al. "Machine Learning in Requirements Engineering: A Mapping Study." 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW), IEEE, 2021, pp. 116–25. DOI.org (Crossref), https://doi.org/10.1109/REW53955.2021.00023.

[3] GITHUB - stpa-step1-dataset. July, 2024. Available at: https://github.com/andreyokamura-unicamp/stpa-step1-dataset.

Figures:
https://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/
https://psas.scripts.mit.edu/home/wp-content/uploads/2023/2023-06-06-1200__Bryan-Smith__PUB.pdf

21 Thank You!

“stpa-step1-dataset” GitHub Repository:

Andrey Toshiro Okamura

Contact:

a213119@dac.unicamp.br